Sound Ontologies. Methods and Approaches for the Description of Sound

11

Sound Ontologies

Methods and Approaches for the Description of Sound

Davide Andrea Mauro, and Andrea Valle

11.1. Introduction

Categorizing and classifying sounds is a useful tool to organize a palette for sound designers and composers. If doing this for acoustic and traditional instruments is a relatively well established task, e.g. Von Hornbostel and Sachs (1961), the same cannot be said for any sound in general. With sounds that are not originated by traditional instruments (or even produced by using traditional instruments in nontraditional ways), we might lack the same rigorous set of criteria, e.g. the originating mechanism: vibrating strings or resonating membranes. For this reason, employing the same classification rules will not necessarily lead to the desired outcomes.

Categories and ontologies can be used to relax the strict requirements of a formal classification and allow users to organize their personal sonic space. As the color palette prepared by a painter contains only a subset of all the possible colors, the tools that we use to synthesize sounds present us with a limited set of options, so rather than being neutral or agnostic tools, they are contributing in shaping our own creativity. Furthermore, ontological representations of sounds are required in order to support a semantic retrieval of sound resources, as in accessing sounds from a library.

Our goal is to present notable attempts in these directions highlighting both the technical/technological substrate and their philosophical approach.

11.2. Phenomenological Approaches

Phenomenological approaches to sound description (e.g. Erickson, 1975) are intended to elicit categories that are perceptually relevant to the listener without an explicit reference to their acoustic properties. While such

categories may be culturally biased, still a careful definition can provide useful ways to identify and describe a variety of sounds. In relation to such an approach, the most relevant proposal is the one by Schaeffer (2017), dating back to 1966. Schaeffer has proposed a double-sided analytic device—a typo-morphology—intended as a multifaceted tool for the description of all the objects of the audible domain (*sound objects*). In particular, the typology is meant as the description of a sound object in relation to other objects, while the morphology is intended as a description of the sound object per se. Starting from the latter, morphological criteria are defined as a set of seven analytical properties (i.e. parameters having different values) characterizing a sound object. These criteria are (Chion, 2009):

1. *Mass*—mode of occupation of the pitch-field by the sound. Differently from pitch, mass takes into account two notions: site as a position on the continuum (i.e. as the actual register of the sound object) and caliber, indicating properly a range of occupation. Pitched sounds thus have a limited caliber, while noisy sounds have a greater caliber.
2. *Harmonic timbre*—diffuse halos of the sound and associated qualities that seem to be linked with mass
3. *Dynamic*—development of sound in the intensity-field
4. *Grain*—micro-structure of the matter of the sound, suggesting the texture of a cloth or mineral
5. *Allure*—oscillation, characteristic vibrato of the sustainment of sound
6. *Melodic profile*—general profile of a sound developing in tessitura
7. *Mass profile*—general profile of a sound where the mass is sculpted by internal variations

Taken together, these criteria are able to describe in detail many qualitative aspects of a sound. The morphological point of view has been widely reconsidered by Smalley (1986 and 1997), who has proposed a *spectromorphology*. The term clearly refers to spectral content of sound, but the proposal does not take into account physical notions or measures. In a spectral typology, the continuum *note-noise* includes *node* as its middle term. Spectral content can be basically categorized along the axis *gesture-texture*. In relation to sound, gesture indicates a figurative bonding toward an acoustic model and a clear direction in development. Texture is related to internal behavior patterning. Sound gestures are described according to three *morphological archetypes* that can be combined into *morphological models*: *attack*, *attack-decay*, *graduated continuant*. Spectral motion, i.e. the way in which sound evolves, can be organized into a typology based on five basic types: *unidirectional*, *bidirectional*, *reciprocal*, *centric/cyclic*, *eccentric/multidirectional*. At a higher level (i.e. in relation to complex, evolving sounds), Smalley (1986) proposes a classification of 15 structural

functions, modeled following the morphological archetypes. An example of description (related to spectral density) is shown in Figure 11.1.

Differently from morphology's analytical criteria, Schaeffer (2017)'s typology is meant as a way to define each sound object in relation to other sound objects. Six typological categories for sound description are identified (*mass, variation, duration, sustain, facture, balance*), then they are tentatively combined in a two-dimensional space for sake of simplicity. This space is a sort of cartography of potential sounds (Risset, 1999), and each object can be described by assigning it to a position. In its final arrangement, the typological space is divided in 28 areas, representing typological labeled *classes*, and the areas are grouped into three *regions* (balanced, slightly original, too original). Every sound (object) thus belongs to a certain class and consequently to a certain region.

Valle (2015 and 2016) has suggested a simplified revision of Schaeffer's space by isolating four (rather than six) categories: *sustain, profile, mass, variation*. Sustain describes a sound object's internal temporality. Thus, in relation to sustain, it is possible to individuate three cases:

1. *Sustained*—constant activity over time
2. *Impulsive*—activity as a singular moment
3. *Iterative*—activity as a series of repeated contributions

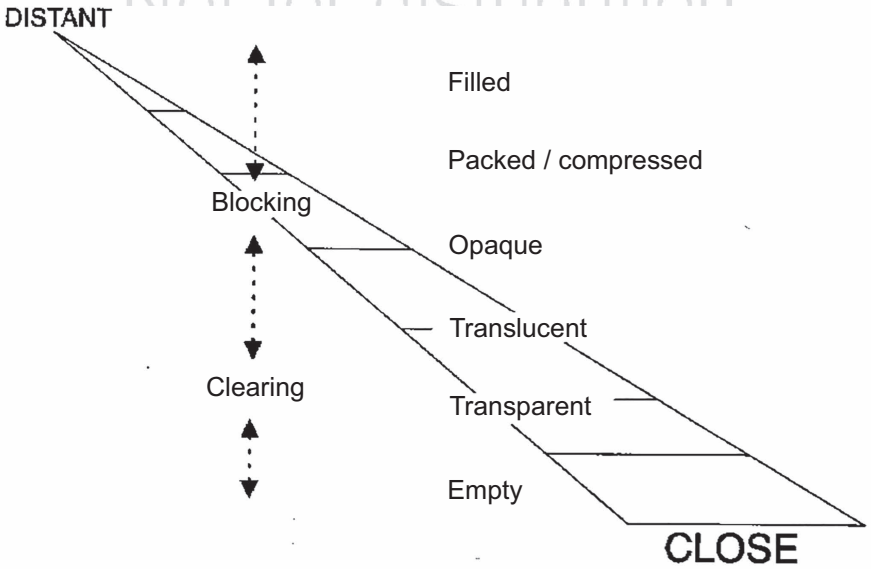

Figure 11.1 Description of Spectral Density.

Source: From Smalley (1997)

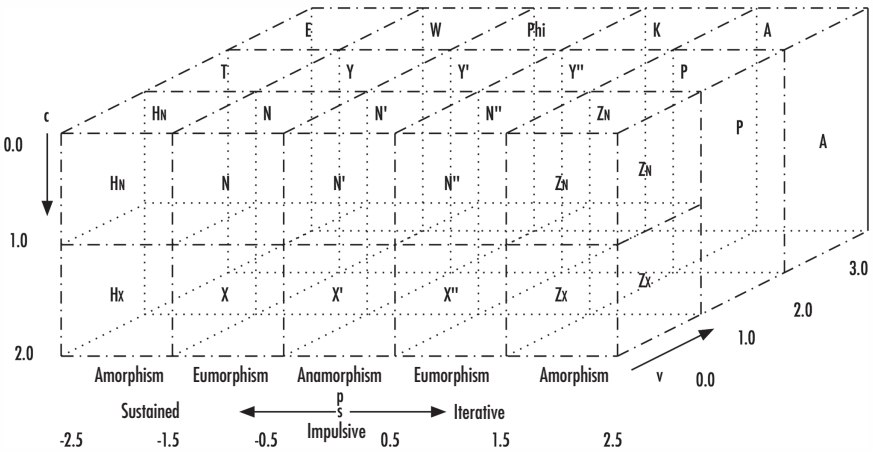

Figure 11.2 Typological Space for Sound Objects.

Source: From Valle (2015)

While *sustain* defines the way in which a sound object is maintained into duration, *profile* describes its external temporal form, in relation to beginning, duration, ending:

1. *Eumorphism*—relevance of all the three categories. The sound object has a well-defined temporal shape.
2. *Amorphism*—Duration is relevant, while beginning and end are not (amorphous sounds lasts indefinitely or do not depend on beginning/end).
3. *Anamorphism*—Profile is compressed, and duration is not relevant (sound objects as events).

Sustain and *profile* are orthogonal categories that collapse in the case of impulsive sustain and amorphous profile. Finally, temporality in specific relation to mass is articulated by Schaeffer by introducing *variation* as a criterion, which allows us to describe how much the mass (site/caliber) changes in time (from stable to varying objects). The previous dimensions can be combined into a three-dimensional space (Figure 11.2), where letters represent classes referring to Schaeffer's usage, and the axes receive arbitrary numerical ranges that have the only means of providing a reference for an explicit annotation.

11.3. Voice-Based Approaches

The qualitative features of sound may find a reference in acoustic instruments and everyday objects. In this sense, the human voice provides a

basic, embodied tool against which to describe and categorize sounds. *Articulatory phonetics* deals with the descriptions of the mechanics of sound production in speech. While indeed not all sounds are available for human vocal production, a wide variety is in use in languages and can act as a reference for sound identification. The International Phonetic Association (IPA) has proposed over the years an International Phonetic Alphabet (also IPA) to annotate speech sound.[1] The main distinction is between consonants (unpitched, noisy) and vowels (pitched). In the IPA, consonants are organized in a two-dimensional chart (Figure 11.3) by their *place* (which part of the vocal tract is obstructed) and *manner* (how it is obstructed).

The resulting chart allows for the description and annotation of many sounds and for the possible identification of their similarity. Vowels are described by IPA by means of a space that couples *height* and *backness* (Figure 11.4).

The first is related to the aperture of the jaw (close-open); the second indicates the position of the tongue relative to the back of the mouth. Such a space is continuous and may provide the sound designer hints to describe harmonic, pitched sounds in terms of vowel qualities. As an example, Takada et al. (2010) use IPA transcriptions to annotate environmental sounds. The study of speech has prompted other general investigations on the description of sound qualities. Since Jakobson et al. (1952), acoustic phonetics has been instrumental in exploiting sonograms as compact time/frequency representations for sounds and in proposing descriptive categories to differentiate spectral mixtures. Moving from such studies, Cogan (1984) has proposed 13 categories for the general interpretation of spectral phenomena (Figure 11.5). In the annotation of sound spectra, these categories can receive four different values: negative, positive (if respectively the first or the second one is dominant), mixed (if both are present) or neutral (if not relevant). It can be observed that both frequency content and time are taken into account.

CONSONANTS (PULMONIC) © 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

Figure 11.3 IPA alphabet: Consonants.

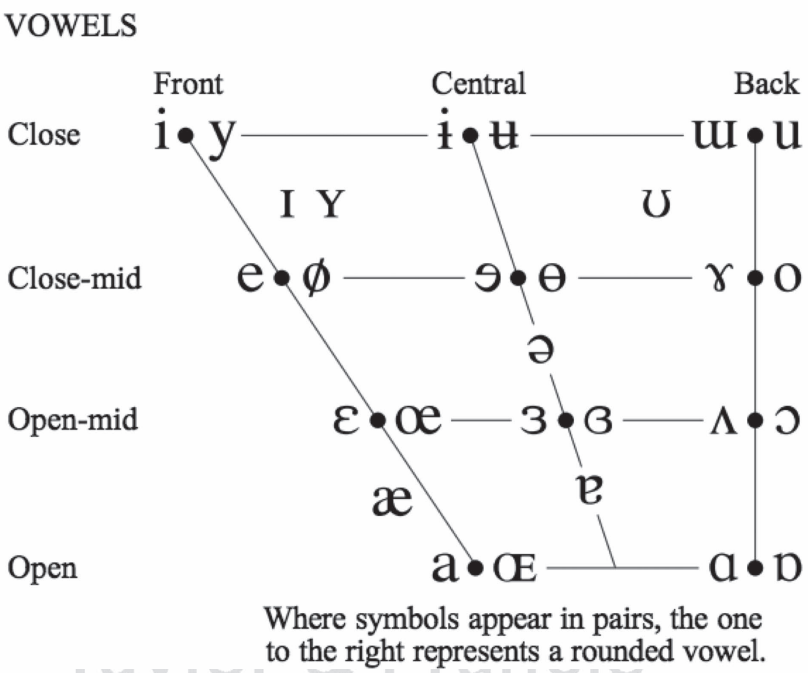

Figure 11.4 IPA Alphabet, Vowels.

On the same path, Slawson (1981, 1985) explicitly distinguishes timbre from sound color. While in Cogan the temporal dimension is still relevant, in Slawson sound color is what remains of sound qualities once time features are eliminated. Starting from phonological categories by Jakobson et al. (1952) and Jakobson and Halle (1956), Slawson (1985) has proposed a two-dimensional space for a specific subset of timbre, named *sound color*, and inspired by vowel formant space. A formant space is constructed by coupling on its two axes the frequencies of the first two formants, i.e. spectral peaks of a vowel (typically indicated as F1 and F2). Such an organization is able to provide a clear definition of vowels in terms of positions into specific regions of the formant space (Fant, 1960). Starting from such a space, Slawson hypothesizes three dimensions for sound color: *openness*, *acuteness*, *laxness*. They are correlated to the physical characteristics of the speech filter through the two formant frequencies. Openness indicates the opening of the oral cavity as a filter deformation ([i]–[æ]); acuteness grows according to the second resonance ([u]–[i]); laxness indicates the state of relaxation of muscle tension ([u]–[ə], a vowel produced by no constriction at all). A fourth dimension, *smallness*, models the length of the vowel tube and can be thought as the height involved in the vowel (resulting from the difference between the two formants, [u]–[a]).

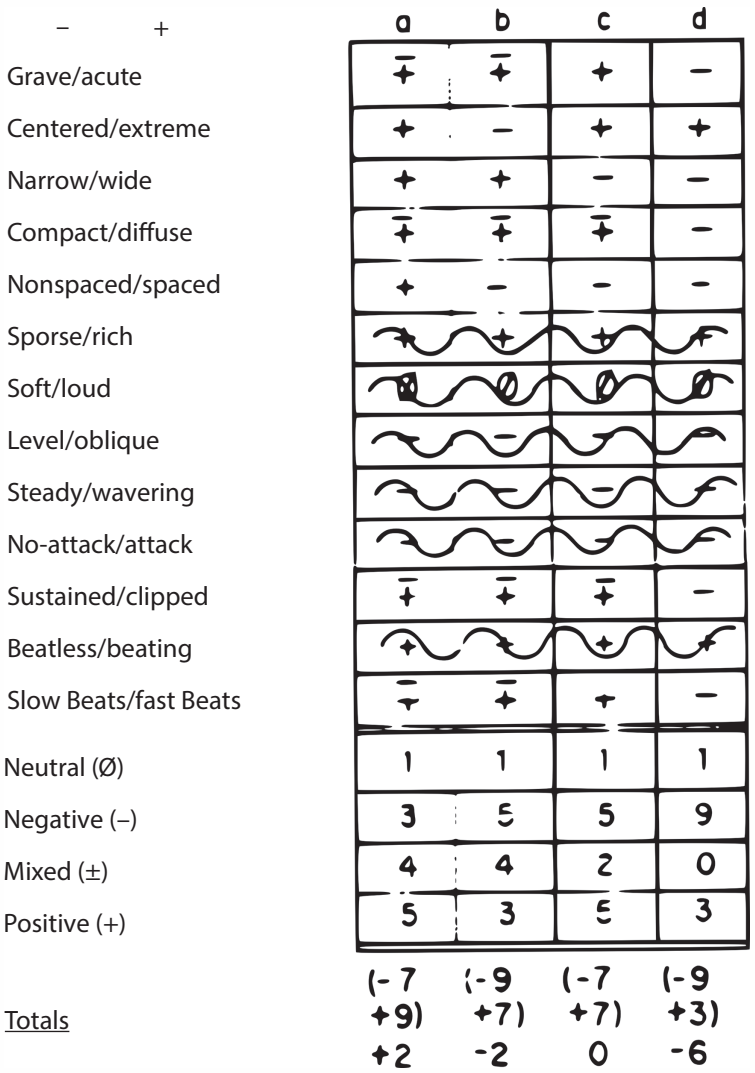

Figure 11.5 An Example of Analysis Chart by Annotation of Spectral Categories.

Source: From Cogan (1984)

Precisely as a consequence of vocal abstraction, the formant space F1/F2 is a strictly continuous one. For each of the four dimensions, Slawson defines a family of *loci* in which the value of the dimension is invariant, i.e. isometric contours of equal openness, equal acuteness, equal laxness and equal smallness (Figure 11.6, in which the author uses "ne" for [Ɵ]; equal smallness is omitted). A relevant feature of this space is that, even

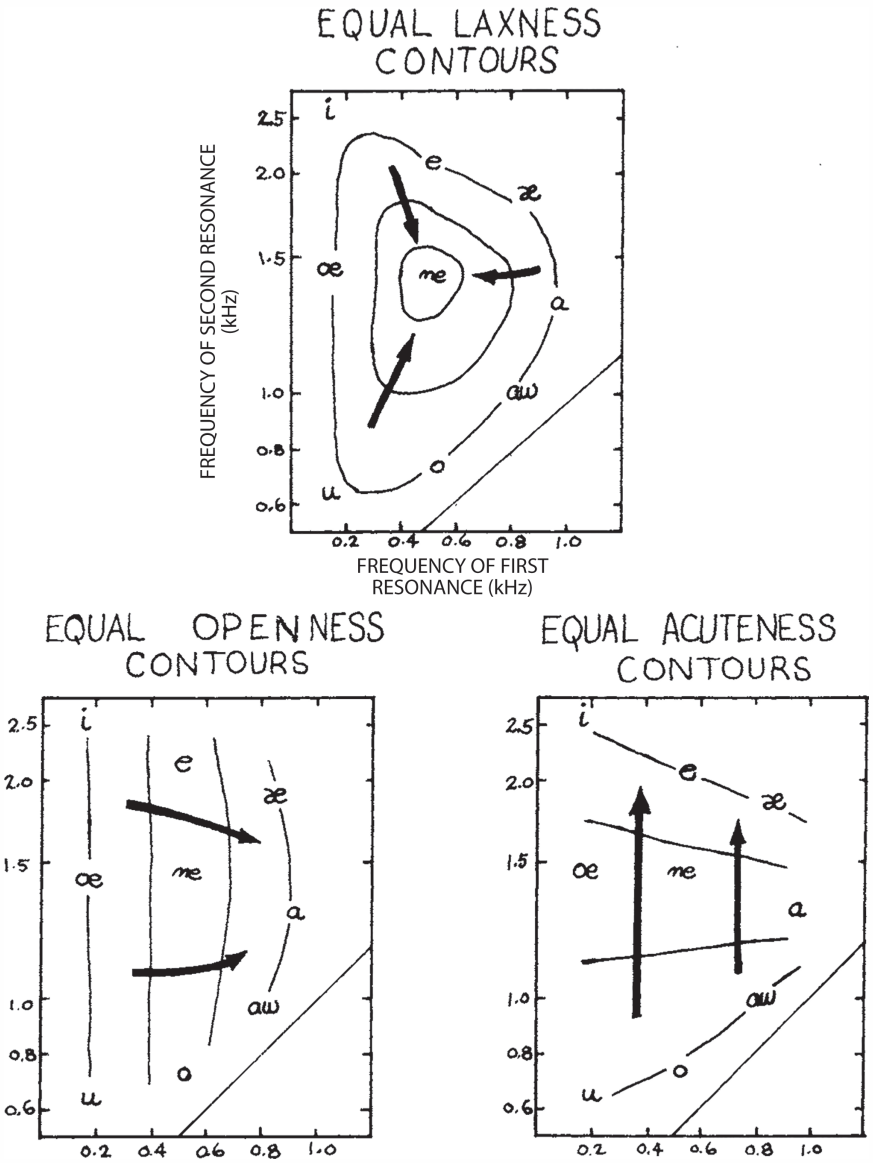

Figure 11.6 Three Dimensions of Sound Color in a F1/F2 Space.

Source: From Slawson (1981)

if inspired from acoustic measurements of formant frequencies, in the end it refers to the speech articulation. In short, a sound color is defined as a vowel color. In this sense, Slawson's space may be thought of also as a phenomenological one, as timbre can be described with reference to the human voice. An operational aspect of the space is that it allows timbre transpositions as geometrical translations. Along the same path, McAdams and Saariaho (1991) have proposed a voice-based organization of timbre.

11.4. Psychoacoustic Approaches

Classification of sounds has been pursued in psychoacoustic studies mainly in relation to *timbre*. The latter is intended as a qualitative feature of sound, i.e. the subjective counterpart of the spectral composition of tones, even if it has been proved that the temporal behavior crucially contributes to such a qualitative assessment (Rasch and Plomp, 1982).

Timbre cannot be ordered on a single scale, as it is a multidimensional attribute of the perception of sound. Hence the need to identify various attributes and to arrange them via multidimensional scaling. A common technique is to collect similarity judgments (Plomp, 1976; Rasch and Plomp, 1982; Grey, 1977; Wessel, 1979) and to arrange them into a space (a *timbral space*), in which geometry respects their similarities. The arrangement per se does not provide semantic categories. Yet, as many times the sounds are collected from acoustic instruments (e.g. organ stops in Plomp, 1976 and Rasch and Plomp, 1982 or orchestral instruments in Grey, 1977 and Wessel, 1979), the reference to commonly used musical instruments may act as reference for sound identity. The visual representation of timbral spaces provides per se a sort of similarity map that can be explored and exploited in the production context. This is explicitly the aim of Wessel (1979) in defining a parallelogram model that is able to predict timbre analogies by means of geometrical patterns in the space, to be used in sound design (analogously to what happens in Slawson, as previously explained). In timbre studies, many semantic categories (i.e. couples of verbal terms intended to be opposite) have been proposed to categorize sounds, to be possibly matched with (hence causally motivated by) acoustic features. As an example, Bismarck (1974) provides to listeners 30 verbal categories (like hard-soft, sharp-dull, coarse-fine). While such categories are proposed in input to the listeners, other verbal categories result from the interpretation of timbral spaces, mostly by individuating common acoustic features (Handel, 1989). Bismarck (1974) sums up his research proposing *sharpness* and *compactness*. The first is related to the distribution of spectral energy around the higher-frequency region, the second is a factor distinguishing between tonal (compact) and noisy

(noncompact) aspects of sound (Rasch and Plomp, 1982). Plomp (1976) suggests few versus many strong higher harmonics for his space. Poli and Prandoni (1997) indicate *brightness* (boosting of the fundamental frequency) and *presence* (spectrum midband enhancement). The three axes in Grey (1977)'s space (Figure 11.7) can be interpreted as spectral energy distribution (narrow-wide), synchronicity (in the collective attacks and decay of upper harmonics, i.e. spectrally stable versus fluctuating), attack dispersion (high-frequency, scattered energy versus energy concentrated on the fundamental frequency, i.e. buzz-like versus soft attack).

It is apparent how, together with spectral features, the temporal dimension (i.e. the attack) is a crucial factor in identifying sound. Accordingly, Wessel (1979) describes the two axes of his two-dimensional arrangement (Figure 11.8) in relation to the spectral energy distribution of the tones and to the nature of the onset transient. The resulting categories can be

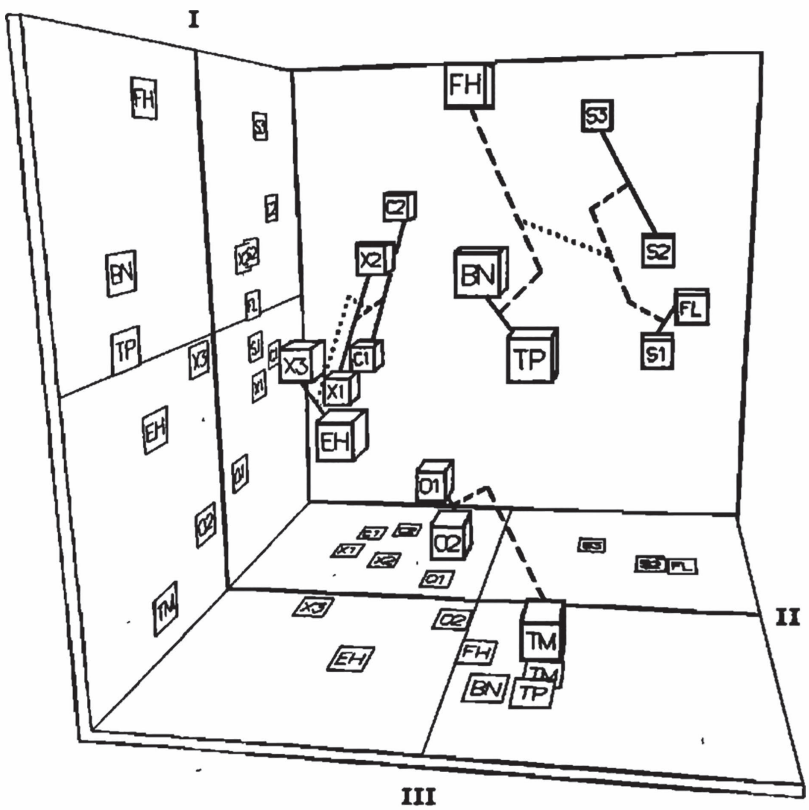

Figure 11.7 Three-Dimensional Spatial Solution for 35 Sounds.

Source: From Grey (1977)

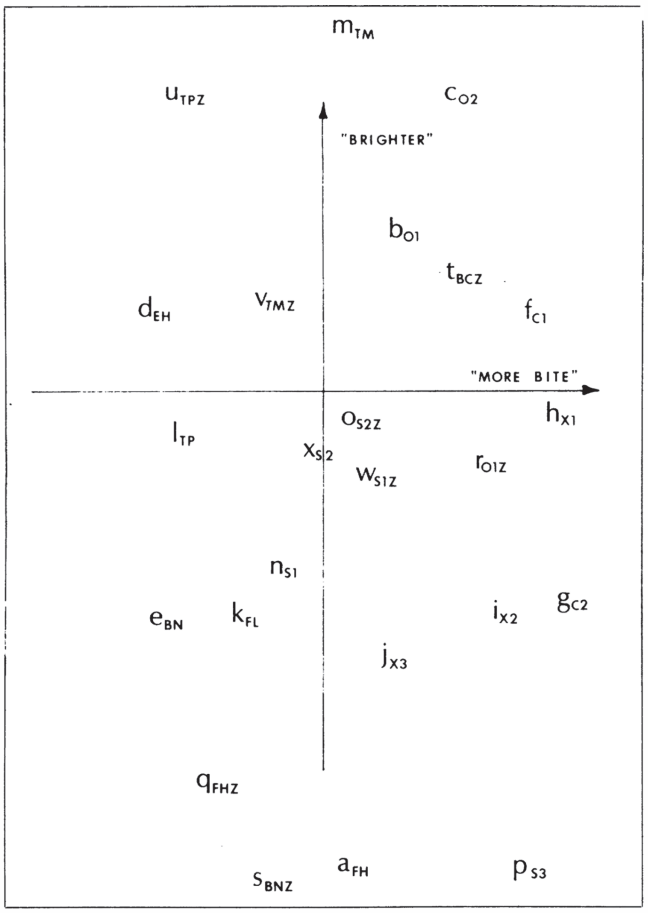

Figure 11.8 Two-Dimensional Timbre Space Representation of 24 Instrument-Like Sounds.

Source: From Wessel (1979)

labeled, respectively, as *bright/mellow* and *soft/biting*. McAdams (1999) individuates three dimensions. Spectral centroid is the center of gravity of the spectrum, spectral flux is intended as the degree of variation of spectrum in time, and log attack time is indeed related to onset time. To these, spectral smoothness can be added, as the degree of amplitude variation between adjacent partials.

To sum up, various categories have been proposed to describe timbre as an overall quality of sounds.

While temporal aspects have proven to be relevant, still it seems a general feature can be found that can be extracted from most of the previous

discussion and seems to be a semantic correlative of an energetic acoustic quality. *Brightness* (versus *dullness*) is thus the one dimension that tends to be found across a variety of studies (Bregman, 1990).

11.5. Ecological Approaches

Ecological approaches to sound have introduced specific ways to describe sounds in relation to their environmental context. The *soundscape* is thus the sonic counterpart of the landscape.

In his foundational work, Murray Schafer (1977) introduced the lo-/hi-fi categorization and a tripartite classification of sound material into *keynote sounds*, *sound signals*, and *soundmarks*. Starting from the former, keynote sounds are the sounds heard by a particular society continuously or frequently enough to form a background against which other sounds are perceived (e.g. the sound of the sea for a maritime community). Signals stand to keynote sounds as a figure stands to a background: they emerge as isolated sounds against a keynote background (e.g. a fire alarm). Soundmarks are socially or historically relevant signals (e.g. the ringing of the historical bell tower of a city).

In relation to soundscape, Böhme (2000) has proposed an aesthetics of *atmospheres*. Every soundscape has indeed a specific scenic atmosphere, which includes explicitly an emotional and cultural dimension. An atmosphere is an overall layer of sound that cannot be analytically decomposed into single sound objects, as no particular sound object emerges from it. While keynote sounds are intended as background sounds (i.e. they are a layer of the soundscape), atmospheres identify the whole sound complex.

The *lo-/hi-fi* categorization differentiates soundscape in relation to the presence of large masking sounds. It is both a theoretical and a historical classification. Premodern soundscape, where low-intensity sonic details were audible, is typically hi-fi, while modern soundscape, after the electromechanic revolution, is typically lo-fi. Apart from this historical context, the distinction is useful as a general classification of soundscapes.

Krause (2015) has extensively used the tripartition *geophony*/*biophony*/*anthrophony* to characterize environmental sounds in relation to three layers of production, respectively natural nonliving phenomena, living beings and humans.

11.6. Analytical Approaches

It is important to note that, rather than a single unique classification approach to sound, many possible perspectives can emerge, each one highlighting a

specific mechanism behind it (e.g. what types of similarities are used) and giving birth to a peculiar set of categories.

Houix et al. (2012) present experiments aimed at classifying environmental sounds and strategies for their categorization. Their analysis starts with the first attempt of Vanderveer (1979) that shows participants grouping sounds together by analyzing the cause that produced such interaction or according to some acoustic properties.

Marcell et al. (2000) report a classification of 120 environmental sounds within 27 very heterogeneous categories corresponding to sound sources (e.g. four-legged animal, air transportation, human, tool, water/liquid), locations or contexts (kitchen, bathroom) or more abstract concepts (hygiene, sickness).

Gygi et al. (2007) report similar results, finding 13 major categories based on 50 sounds. The most frequently used categories referred to the type of sources (e.g. animals/people, vehicles/mechanical, musical and water). In a lesser proportion, sounds were grouped by context (e.g. outdoor sports) or location (e.g. household, office, bar).

Guyot et al. (1997) propose a framework for the classification of environmental sounds based on two strategies: the first is based on psychoacoustic criteria (e.g. pitch, temporal evolution), while the second is based on the identification of the source. They use the three levels of abstraction formalized by Rosch and Lloyd (1978): superordinate, base and subordinate levels. At the superordinate level, listeners identify the abstract mechanism of sound production. At the base level, they identify actions. And at the subordinate level, they identify the source. The different types of categories are not mutually exclusive and can be mixed during a classification task (across participants and/or for a single participant) because a sound can belong to multiple categories corresponding to different conceptual organizations. This cognitive process has been called "cross-classifications" in Ross and Murphy (1999).

A notable attempt at classifying everyday sounds has been proposed by Gaver (1993a, 1993b) representing different classes of physical interactions (solids, liquids, gases). The system has a hierarchical structure (similar to a taxonomy) and is based on the physics of sound-producing events. Gaver himself observes that the framework is not exhaustive and that entirely different ways of organizing the materials are indeed possible.

To read more on this topic, see also Stefano Delle Monache and Davide Rocchesso's chapter, "Sketching Sonic Interactions," in the third volume of this series, *Foundations in Sound Design for Embedded Media*, where the authors present a way to organize sonic material to build a personalized sonic sketchbook.

A common problem with classification is that different users, even adopting the same taxonomy, can classify the same object into different

classes, and they might want to do that assigning a different "degree of membership." In Ferrara et al. (2006), the authors investigate an ontology aimed at describing music pieces and address this specific problem in terms of genre classification. Assigning a degree of membership is typical of fuzzy logics, and different strategies can be employed to modify standard tools to support the aforementioned concept. The problem can be addressed either by extending the tools to fully support those concepts at the cost of losing compliance and support of the main standard tools used for working with ontologies or by finding ways of expressing the same concepts from within the standard. For the latter solution, extra work is required in order to tweak the tools at the risk of losing a bit of simplicity but retaining compliance with the standards.

An important contribution to the definition of sound ontologies comes from the field of computational auditory scene analysis (CASA) where the definition of suitable ontologies is a requirement in order to provide a meaningful classification of the events. In Nakatani and Okuno (1998), the sound ontology is composed of three elements: *sound classes*, definitions of *individual sound attributes*, and their *relationships*. The ontology is defined hierarchically by using:

1. *Part-of*—a hierarchy based on the inclusion relation between sounds.
2. *Is-a*—a hierarchy based on the abstraction level of sound.

The Part-of hierarchy of basic sound classes is composed of four layers of sound classes. A *sound source* is a temporal sequence of sounds generated by a single sound source. A *sound source group* is a set of sound sources that share some common characteristics as music. A *single tone* is a sound that continues without any durations of silence, and it has some low-level attributes such as harmonic structure. In each layer, an upper class is composed of lower classes that are components sharing some common characteristics. For example, a harmonic stream is composed of frequency components that have harmonic relationships. The Is-a hierarchy can be constructed using any abstraction level. For example, voice, female voice, the voice of a particular woman and the woman's nasal voice form make up an Is-a hierarchy. With sound ontology, each class has some attributes, such as fundamental frequency, rhythm and timbre. A lower class in the Is-a hierarchy inherits the attributes of its upper classes by default. In other words, an abstract sound class has attributes that are common to more concrete sound classes.

Burger et al. (2012), without explicitly referring to ontologies, define 42 "noisemes" as fundamental atomic units of sound capturing objective properties of the acoustic signal. The labels are used in a classification task for environmental noise sounds.

Fields (2007) reflects on the difference between a top-down approach (ontology), and a bottom-up approach (folksonomies) and applies this conceptualization framework to building an audio ontology describing the *Computer Music Tutorial* by Curtis Roads using the Protégé tool (Musen, 2015).

In Lobanova et al. (2007), the authors deal with sounds that cannot be categorized by linking them to a source concept. Their implementation is based on WordNet (Miller, 1995), a widely used lexical resource in computational linguistics.

In Bones et al. (2018), the authors emphasize how categorization of sounds is based upon different strategies depending on context and the availability of cues. The study is focused on the categorization of three different types of environmental sound: dog, engine and water sounds, for which subjects were able to describe sounds with three types of attributes: the *source-event* (referring to the inferred source of the sound), the *acoustic signal* (explicitly referring to the sound itself) or a *subjective-state* (describing an emotional response caused by the sound or the sound source).

11.7. Technical Tools and Applications

In order to semantically enhance the retrieval of sound files (or profiles), many have attempted to define an annotation and classification schema. The first technical problem is the language of such scheme. With the advent of semantic technologies for the web and RDF-based vocabularies, semantic interoperability has become one of the desiderata of data models. In this regard, a tendency that showed up is the confluence of different domain-specific vocabularies in more general ontologies used for data integration. In 2012, the World Wide Web Consortium (W3C) mapped some of the most used schemas for media objects in the Ontology for Media Resources: OWL Web Ontology Language (VV.AA., 2012), recommending it for the annotation of digital media on the web. The W3C specification is not bound to a serialization in a particular language, so it can be used as a general schema.

In Hatala et al. (2004), the authors use ontologies to retrieve sound objects in an augmented reality (AR) application for museums. This implementation uses DAML+OIL (a standard now superseded by OWL) but highlights how the designers valued the possibility of performing reasoning with the system to automatically retrieve digital objects. The auditory interface follows an ecological approach to sound composition. Three areas are taken into account: psychoacoustic, cognitive and compositional. Psychoacoustic features of the ecological balance include spectral

balancing of audible layers. Cognitive aspects of listening are represented by content-based criteria. Compositional aspects are addressed in the form of the orchestration of an ambient informational soundscape of immersion and flow that allows for the interactive involvement of the visitor.

In the SoDA project, Valle et al. (2014) organize a collection of sounds and implement a semantic search engine based on classical techniques borrowed from information retrieval (IR), whose main task is finding relevant "documents" on the basis of the user's information needs, expressed to the system by a query. The annotation is based on an OWL schema inspired by annotation in state-of-the-art sound libraries. Among the libraries used by sound designers taken into consideration are Sound Ideas Series (6000, 7000 and 10000), World Series of Sound, Renaissance SFX. The search tools for audio documents taken into account were SoundMiner,[2] Library Monkey,[3] Basehead,[4] Audiofinder.[5]

Audio Set, presented in Gemmeke et al. (2017), is a large-scale data set of manually annotated audio events using a structured hierarchical ontology of 632 audio classes. One of the aims of the authors is to bridge the relatively large gap that still exists between image recognition and sound recognition, providing a comprehensive coverage of real-world sound. The ontology is released as a JSON file.[6] In the creation of the ontology, in order to avoid biasing the categories, the authors started from a neutral, large-scale analysis of web text.

Even if not formally presented as ontologies, a number of tools have attempted to organize a sonic space in order to enable the composer/performer to search and act on such a potentially vast space.

In Rocchesso et al. (2016), the authors propose to represent the sonic space of a sound model as a plane where a number of prototype synthetic sounds are positioned. The spatial organization is based on a dimensionality reduction on the set of available sound, each represented by a high-dimensional feature vector. Two-dimensional spaces are particularly relevant for sound designers (see the preceding discussion) because they can be used as sonic maps, possibly accompanied by few landmarks that are highlighted and serve the role of prototypical sounds for a certain "class."

For similar tasks, the first problem is how to represent and describe our sounds: whereas digital signals are described by sequences of many values, we want to obtain compact descriptions that can be better manipulated. In the area of music information retrieval, a lot of research has been devoted to automatically extract descriptors (or features; the MIRToolbox presented in Lartillot and Toiviainen 2007 is widely adopted for these tasks) that could concisely represent sounds. Once the sounds are associated with a compact representation, it is possible to try to organize them in a low-dimensional space (typically two or three). A classic way to do

that is by means of principal component analysis (PCA), which is based on singular value decomposition (SVD). An in-depth description of all these methodologies is beyond the scope of this chapter, but Drioli et al. (2009), Scavone et al. (2001), and Fernström and Brazil (2001) can be used as starting point to explore how to apply those techniques to this task.

The main goal of the previously discussed tools is to allow the users to use the power of IR techniques to be able to search and navigate through a collection of sounds. It is important to point out that a main limitation is that some of the processes presented here are manual or at least partially supervised by human input. The reason behind this characteristic resides in the already mentioned polymorphic nature of classification strategies.

11.8. Conclusions

In this chapter, we presented an overview on the topic of ontologies and, more in general, the classification and organization of sounds. We immediately realized how it is not possible to define a single unifying approach to sound organization because many different criteria may be useful in the classification tasks depending on various factors. As it is not either possible or useful to define a single classification strategy for sounds, in sound design a multifaceted approach is the most apt to cope with the variety of design contexts. Following these assumptions, we thus presented the reader a number of approaches grouped into five different categories (phenomenological, voice-based, psychoacoustic, ecological and analytical) highlighting the various possible mechanisms that can be employed by the sound designer when she or he wants to attempt a personal classification/organization of a sound materials. In short, the tools that the sound designer seeks to use and develop need to be flexible enough to reflect this variety of contexts and possibilities. The application examples that we introduced in the final section are thus intended to show how to exploit, at various levels, the aforementioned strategies.

Notes

1. www.internationalphoneticassociation.org
2. http://store.soundminer.com
3. www.monkey-tools.com/products/library-monkey/
4. www.baseheadinc.com
5. www.icedaudio.com
6. http://g.co/audioset

References

Böhme, G. (2000). Acoustic atmospheres. A contribution to the study of ecological aesthetics. *Soundscape: The Journal of Acoustic Ecology*, I(1), pp. 14–18.

Bones, O., Cox, T. J. and Davies, W. J. (2018). Distinct categorization strategies for different types of environmental sounds. In: *Euronoise 2018*, Crete, EEA-HELINA.

Bregman, A. (1990). *Auditory scene analysis*: *The perceptual organization of sound*. Cambridge, MA and London: MIT Press.

Burger, S., Jin, Q., Schulam, P. F. and Metze, F. (2012). *Noisemes: Manual annotation of environmental noise in audio streams*. Technical Report CMU-LTI-12–07, Carnegie Mellon University.

Chion, M. (2009). *Guide to sound objects*. Unpublished.

Cogan, R. (1984). *New images of musical sound*. Cambridge, MA: Harvard University Press.

Drioli, C., Polotti, P., Rocchesso, D., Delle Monache, S., Adiloglu, K., Annies, R. and Obermayer, K. (2009). Auditory representations as landmarks in the sound design space. In: *Proceedings of the sixth sound and music computing conference*. Porto, Portugal, pp. 315–320.

Erickson, R. (1975). *Sound structure in music*. Berkeley, Los Angeles and London: University of California Press.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague, Netherlands: Mouton & Co N.V, Publishers.

Fernström, M. and Brazil, E. (2001, July 29–August 1). Sonic browsing: An auditory tool for multimedia asset management. In: J. Hiipakka, N. Zacharov and T. Takala, eds., *Proceedings of the seventh international conference on auditory display (ICAD)*, Espoo, Finland.

Ferrara, A., Ludovico, L. A., Montanelli, S., Castano, S. and Haus, G. (2006). A semantic web ontology for context-based classification and retrieval of music resources. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2(3), pp. 177–198.

Fields, K. (2007). Ontologies, categories, folksonomies: An organised language of sound. *Organised Sound*, 12(2), pp. 101–111.

Gaver, W. W. (1993a). How do we hear in the world? Explorations in ecological acoustics. *Ecological Psychology*, 5(4), pp. 285–313.

Gaver, W. W. (1993b). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5(1), pp. 1–29.

Gemmeke, J. F., Ellis, D. P. W., Freedman, D., Jansen, A., Lawrence, W., Moore, R. C., Plakal, M. and Ritter, M. (2017). Audio set: An ontology and human-labeled dataset for audio events. In: 2017 *IEEE International conference on acoustics, speech and signal processing (ICASSP*), New Orleans, LA, pp. 776–780.

Grey, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am*. 61(5), pp. 1270–1277. Reproduced with the permission of the Acoustical Society of America.

Guyot, F., Castellengo, M. and Fabre, B. (1997). Chapitre 2: étude de la catégorisation d'un corpus de bruits domestiques. In: *Catégorisation et cognition: de la perception au discours*. Paris: Editions Kimé, pp. 41–58.

Gygi, B., Kidd, G. R. and Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics*, 69(6), pp. 839–855.

Handel, S. (1989). *Listening. An introduction to the perception of auditory events*. Cambridge, MA and London: MIT Press.

Hatala, M., Kalantari, L., Wakkary, R. and Newby, K. (2004). Ontology and rule based retrieval of sound objects in augmented audio reality system for museum visitors. In: *Proceedings of the 2004 ACM symposium on applied computing*. Nicosia, Cyprus: ACM, pp. 1045–1050.

Houix, O., Lemaitre, G., Misdariis, N., Susini, P. and Urdapilleta, I. (2012). A lexical analysis of environmental sound categories. *Journal of Experimental Psychology: Applied*, 18(1), pp. 52–80.

Jakobson, R., Fant, C. M. and Halle, M. (1952). *Preliminaries to speech analysis. The distinctive features and their correlates*. Cambridge, MA: MIT Press.

Jakobson, R. and Halle, M. (1956). *Fundamentals of language*. Berlin and New York: Mouton de Gruyter.

Krause, B. (2015). *Voices of the wild: Animal songs, human din, and the call to save natural soundscapes*. New Haven, CT: Yale University Press.

Lartillot, O. and Toiviainen, P. (2007). A MATLAB toolbox for musical feature extraction from audio. In: *Proceedings of the 10th international conference on digital audio effects (DAFx-07)*. Bordeaux, France, pp. 237–244.

Lobanova, A., Spenader, J. and Valkenier, B. (2007). Lexical and perceptual grounding of a sound ontology. In: J. G. Carbonell and J. Siekmann, eds., *International conference on text, speech and dialogue, lecture notes in artificial intelligence*, Vol. 4629. Berlin and Heidelberg: Springer, pp. 180–187.

Marcell, M. M., Borella, D., Greene, M., Kerr, E. and Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, 22(6), pp. 830–864.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), pp. 39–41.

McAdams, S. (1999). Perspectives on the contribution of timbre to musical structure. *Computer Music Journal*, 23(3), pp. 85–102.

McAdams, S. and Saariaho, K. (1991). *Le timbre. Métaphore pour la composition*. Paris: Christian Bourgois-IRCAM. Chapter Qualités et fonctions du timbre musical, pp. 164–180.

Murray Schafer, R. (1977). *The tuning of the world*. New York: Knopf.

Musen, M. A. (2015). The Protégé project: A look back and a look forward. *AI Matters*, 1(4), pp. 4–12.

Nakatani, T. and Okuno, H. G. (1998). Sound ontology for computational auditory scene analysis. In: *Proceeding for the 1998 conference of the American association for artificial intelligence*, AAAI, pp. 1004–1010.

Plomp, R. (1976). *Aspects of tone sensation: A psychophysical study*. London: Academic Press.

Poli, G. D. and Prandoni, P. (1997). Sonological models for timbre characterization. *Journal of New Music Research*, 26(2), pp. 170–197.

Rasch, R. and Plomp, R. (1982). The perception of musical tones. In: Deutsch, D., ed., *The psychology of music*. Orlando, FL: Academic Press, Chapter 1, pp. 1–24.

Risset, J-C. (1999). *Ouïr, entendre, écouter, comprendre après Schaeffer*. Bryn-sur-Marne and Paris: INA-Buchet, Chastel. Chapter Schaeffer, P.: *recherche et création musicalese et radiophoniques*, pp. 153–159.

Rocchesso, D., Mauro, D. A. and Drioli, C. (2016). Organizing a sonic space through vocal imitations. *Journal of the Audio Engineering Society*, 64(7/8), pp. 474–483.

Rosch, E. and Lloyd, B. B. (1978). *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum.

Ross, B. H. and Murphy, G. L. (1999). Food for thought: Cross-classification and category organization in a complex real-world domain. *Cognitive Psychology*, 38(4), pp. 495–553.

Scavone, G., Lakatos, S., Cook, P. and Harbke, C. (2001). Perceptual spaces for sound effects obtained with an interactive similarity rating program. In: *Proceedings of international symposium on musical acoustics*. Perugia, Italy.

Schaeffer, P. (2017). *Treatise on musical objects. An essay across disciplines*. Oakland: University of California Press.

Slawson, W. (1981). The color of sound: A theoretical study in musical timbre. *Music Spectrum*, 3, pp. 132–141.

Slawson, W. (1985). *Sound color*. Berkeley and Los Angeles and London: California University Press.

Smalley, D. (1986). *The language of electroacoustic music*. London: Macmillan, Chapter Spectromorphology and Structuring Processes, pp. 61–93.

Smalley, D. (1997). Spectromorphology: Explaining sound-shapes. *Organised Sound*, 2(2), pp. 107–126.

Takada, M., Fujisawa, N., Obata, F. and Iwamiya, S-I. (2010). Comparisons of auditory impressions and auditory imagery associated with onomatopoeic representation for environmental sounds. *EURASIP Journal on Audio, Speech, and Music Processing*, (1), p. 674248. https://doi.org/10.1155/2010/674248.

Valle, A. (2015). Towards a semiotics of the audible. *Signata*, 6.

Valle, A. (2016). Schaeffer reconsidered: A typological space and its analytic applications. *Analitica—Rivista online di studi musicali* 8. Available at: www.gatm.it/analiticaojs/index.php/analitica/article/view/158 [Accessed September 2018].

Valle, A., Armao, P., Casu, M. and Koutsomichalis, M. (2014). SoDA: A sound design accelerator for the automatic generation of soundscapes from an ontologically annotated sound library. In: *Proceedings ICMC-SMC-2014,* Athens, Greece, ICMA, pp. 1610–1617.

Vanderveer, N. J. (1979). *Ecological acoustics: Human perception of environmental sounds*. PhD thesis, Cornell University.

von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30, pp. 146–159.

Von Hornbostel, E. M. and Sachs, C. (1961). Classification of musical instruments. Translated from the original German by Baines, A. and Wachsmann, K. P. *The Galpin Society Journal*, pp. 3–29.

VV.AA. (2012). *OWL 2 web ontology language document overview*. Available at: https://www.w3.org/2012/pdf/REC-owl2-overview-20121211.pdf [Accessed March 2019].

Wessel, D. (1979). Timbre space as a musical control structure. *Computer Music Journal* 3, no. 2, pp. 45–52.

W3C. Available at: www.w3.org/TR/2012/REC-owl2-overview-20121211/ [Accessed September 2018].